What is data justice

Taylor, Linnet

Published in:
Big Data & Society

DOI:
10.1177/2053951717736335

Publication date:
2017

Original Research Article

BIG DATA & SOCIETY

Big Data & Society
July–December 2017: 1–14

DOI: 10.1177/2053951717736335
journals.sagepub.com/home/bds
SAGE

What is data justice? The case for connecting digital rights and freedoms globally

Linnet Taylor

Abstract

The increasing availability of digital data reflecting economic and human development, and in particular the availability of data emitted as a by-product of people's use of technological devices and services, has both political and practical implications for the way people are seen and treated by the state and by the private sector. Yet the data revolution is so far primarily a technical one: the power of data to sort, categorise and intervene has not yet been explicitly connected to a social justice agenda by the agencies and authorities involved. Meanwhile, although data-driven discrimination is advancing at a similar pace to data processing technologies, awareness and mechanisms for combating it are not. This paper posits that just as an idea of justice is needed in order to establish the rule of law, an idea of *data justice* – fairness in the way people are made visible, represented and treated as a result of their production of digital data – is necessary to determine ethical paths through a datafying world. Bringing together the emerging scholarly perspectives on this topic, I propose three pillars as the basis of a notion of international data justice: (in)visibility, (dis)engagement with technology and antidiscrimination. These pillars integrate positive with negative rights and freedoms, and by doing so challenge both the basis of current data protection regulations and the growing assumption that being visible through the data we emit is part of the contemporary social contract.

Keywords

Privacy, ethics, development, discrimination, representation, surveillance

Introduction: The case for data justice

As digital data become available on populations that were previously digitally invisible, policymakers and researchers worldwide are taking advantage of what the UN has termed the 'data revolution' (United Nations, 2014). The increasing availability of digital data reflecting economic and human development, and in particular of 'data fumes' (Thatcher, 2014) – data produced as a by-product of people's use of technological devices and services – is driving a shift in policymaking worldwide from being data informed to being data driven (Kitchin, 2016). These granular data sources which allow researchers to infer people's movements, activities and behaviour have ethical, political and practical implications for the way people are seen and treated by the state and by the private sector (and, importantly, by both acting in combination). This distributed visibility has even clearer social and political implications in the case of low-income environments, where authorities' ability to gather accurate statistical data has previously been limited. Yet the data revolution is so far primarily a technical one: the power of data to sort, categorise and intervene has not yet been explicitly connected to a social justice agenda by those agencies and authorities who collect, manage and use data. Nor is there a high level of awareness amongst implementers of how new data technologies may not be neutral in terms of access, use or impacts, something that the available research into this phenomenon shows

Tilburg Institute for Law, Technology and Society (TILT), Netherlands

Corresponding author:
Linnet Taylor, Tilburg Institute for Law, Technology and Society (TILT), P.O. Box 90153, 5000 LE Tilburg, Netherlands.
Email: l.e.m.taylor@uvt.nl

to be the case (Dalton et al., 2016). In fact, while data-driven discrimination is advancing at a similar pace to data processing technologies, awareness and mechanisms for combating it are not.

Two trends make developing a global perspective on the just use of digital data urgently necessary: one is the exponential rise in technology adoption worldwide, and the other the corresponding globalisation of data analytics. Of the world's seven billion mobile phones, 5.5 billion are in low- and middle-income countries (LMICs), where 2.1 billion people are also online (ITU, 2015). India and China have commissioned the creation of hundreds of smart cities that will provide the ability to track and monitor citizens in every aspect of their lives (Greenfield, 2013), digital and biometric registration are becoming the new norm in even the poorest countries, and practices in international aid, development and humanitarian response increasingly use vast amounts of digital data to map, sort and intervene on the mass scale in lower income regions (Taylor, 2015). The reach of the global data market has also changed to take account of these new sources of data, with multinational corporations scrambling to profile billions of potential new consumers (Taylor, 2016a). Meanwhile, research and praxis on the ways in which datafication can serve citizenship, freedom and social justice are minimal in comparison to corporations and states' ability to use data to intervene and influence.

This paper posits that just as an idea of justice is needed in order to establish the rule of law, an idea of *data justice* is necessary to determine ethical paths through a datafying world. Several framings of data justice are emerging within different fields, which have the potential to build on each other. I will therefore analyse the existing work on data justice and place the different viewpoints in dialogue with each other, then argue that by finding common principles we can bring them together into a single framing for further research and debate. The paper is structured as follows: I will first outline the reasons for concern relating to the new public–private interfaces of big data, namely the disciplinary and frequently discriminatory nature of large-scale databases used on the population level, Next, I will use empirical examples to demonstrate that these concerns are not only amplified but fundamentally different in the context of big data. I will then explore current framings of data justice and identify which aspects of the problems arising from big data they propose to address. Next, I will propose an overarching conceptualisation of data justice that can bridge existing approaches and form a basis for dialogue between them. Finally, I will argue for Sen and Nussbaum's Capabilities Approach as a framing for this conceptualisation, with the aim of providing an ecosystemic approach that can address institutions, markets, legal systems and public debates.

A note on methodology: the theoretical and empirical starting points for the framework proposed here are based on a research project comprising fieldwork, observation and interviews conducted over the period 2012–2016. This project included 200 formal and informal interviews and periods of observation conducted with academic researchers, development aid and humanitarian organisations, independent technology developers, activist organisations in the field of data and rights, large technology firms and policymakers from the US, EU and several African and Asian countries. The observation portion of the research was conducted at international events relevant to the 'Responsible Data' movement, through participation in advisory groups and in public discussions on data ethics. The interviews were conducted at these events, and additionally through fieldwork during 2014–2016 at multinational mobile network operators in France and Norway, and on a public-sector datafication project in Bangalore, India.

The problem: Data at the public–private interface

Why look for ways to relate social justice concerns to datafication, and vice versa? Why not, for example, prioritise making sure that commercial digital innovation can proceed unfettered, since this has been argued to benefit everyone in society, or that data fully supports efficiency in the public sector, thus serving the interests of taxpayers and public security? Both of these latter arguments have been made by both high-level private sector (World Economic Forum, 2011) and public-sector actors (European Commission, 2016). What is it about the social impacts of digital data that suggests a social justice agenda is important? For one thing, the impacts of big data are very different depending on one's socio-economic position. The work of Gilliom (2001) and more recently of scholars such as Eubanks (2014) and Masiero (2016) shows that the greatest burden of dataveillance (surveillance using digital methods) has always been borne by the poor. Bureaucratic systems designed to assure that people are not misusing state welfare funds and other publicly funded support are part of the apparatus of governmentality (Lemke, 2001); data-driven law enforcement focuses unequally on poor neighbourhoods which experience certain types of criminality (O'Neil, 2016); and undocumented migrants are tracked and acted upon by digital systems in more invasive ways than higher income travellers (Taylor, 2015).

Beyond socio-economic status, gender, ethnicity and place of origin also help to determine which databases

we are part of, how those systems use our data and the kinds of influence they can have over us. Kang's work on trafficking (2015), for example, shows how the surveillance of women's behaviour and movements by international anti-trafficking and anti-sex work authorities has historically been shaped by very different methodologies and types of expertise depending on subjects' national origin and ethnic group, so that different types of data were fed into the international system from different regions, with corresponding variance in the conceptualisation of who should be the subject of anti-trafficking provisions and of anti-sex-work programmes for control and discipline. Similarly, Moore and Currah's (2015) research on how transgender citizens have been dealt with by population databases in the US shows that one's ability to legally identify as a different gender depends to a great extent on one's income. Jiwani's (2015) work on citizenship and conformity also demonstrates the ways in which surveillance as an 'active social process' reinforces structural and social boundaries.

Moreover, these problems intersect and multiply at the boundaries created by the linking and merging of datasets. This intersectionality (Cho et al., 2013) in the effects of datafication is an important component of the argument for a social justice perspective. A range of interacting characteristics – race, ethnicity, religion, gender, location, nationality, socio-economic status – determine how individuals become administrative and legal subjects through their data and, consequently, how those data can be used to act upon them by policymakers, commercial firms and both in combination. In turn, the possibility of being identified as a target of surveillance multiplies depending on the number of categories of interest one belongs to.

For example, a teenager from an immigrant family, living in a low-income area, whose parents are poor and who belongs to a minority ethnic group and religion is exponentially more likely to be targeted for surveillance by both protective (social services) and preventive at (law enforcement) authorities, and is also likely to have less opportunity to resist that surveillance or intervention than her friend who lives in a high-income area and belongs to the majority ethnic group.

That data systems discriminate is not news. Nor is it news that they tend to further disadvantage those who are already marginalised or socially excluded, or that those people experience the greatest obstacles in seeking redress. The evidence for this is well documented and does not per se argue for a new conceptualisation of data justice – everyone has the right to be treated fairly by public (and private) authorities of all kinds. What does argue for paying special attention to the current implications of datafication for social justice, however, is the particular dynamic of contemporary datafication where methods of data collection and analysis are no longer easily divisible into 'volunteered' (direct surveys or other collection of administrative data, where the citizen is aware her data is being gathered) versus 'other' (digital surveillance via devices and sensors). For the surveilled teenager in the above example, the problem multiplies when the functions of data collection and analysis are shared between public authorities and the commercial firms that provided her phone, her internet access or the apps she uses. The economics of surveillance also have implications for fair representation and access to services, since access to technology increasingly determines who can be seen: Shearmur (2015) has warned that those who use big data to study behaviour or shape policy are seeing not society but 'users and markets'.

The public–private interface is important because many of what we perceive as public-sector functions (counting, categorising and serving our needs as citizens) are in fact performed by the private sector, with corresponding implications for transparency and accountability. The number of public-sector data scientists equipped to analyse big data is tiny in comparison to the number of bureaucrats interested in what big data can tell them, with the consequence that the datafication of government has been, and will always be, executed primarily by the private sector. For example, the whistle-blower Edward Snowden was employed by the consulting firm Booz Allen Hamilton when he performed surveillance for the US intelligence agencies. This suggests that markets are a central factor in establishing and amplifying power asymmetries to do with digital data, and that new strategies and framings may be needed that can address the public–private interface as an important site for determining whether data technologies serve us or control us.

In response to this problem, arguments are emerging within different domains for a broader, social-justice-oriented conceptualisation of our rights with regard to data. As population data become by-products of informational capitalism, this has consequences both for the way we can be monitored and the avenues we have to seek redress if we are subjected to unfair treatment. This is because the tools of law and democratic representation that provide the possibility of redress where personal data is misused become more difficult to use as data starts to flow more freely between commercial and public sector actors. Responsibility and accountability grow fuzzy, partly because each actor can shift the responsibility onto the other, and partly because monitoring is indirect and invisible, making people less likely to identify harms as data related.

The public–private interface involved in large-scale data collection, and its inevitable engagement with the global data market, raise fundamental questions about

how rights can be secured across borders and legal systems, and even about whether individual rights should be the only instrument used to combat data harms (Taylor et al., 2017). One important shift is that the surveillance (or monitoring) relationship, which underpins many other, often positive, functions of data, is no longer one to one with a fixed aim and geography, but rather many to many, virtual and has aims that may shift from governmental to commercial and back again. A panopticon (Foucault, 1977) where continuous surveillance drives people to modulate their behaviour is no longer the most useful metaphor for a contemporary datafied surveillance that is invisible and plural, operating through a myriad different platforms. Instead of censoring our behaviour to please our watchers, we make ourselves accidentally visible through our everyday behaviour to a huge range of actors, from the corporations that make the devices and systems we use, and the service providers who facilitate their content, to the data brokers who track our use of them and the myriad consumers of their products, which include governments, marketing firms, intelligence agencies and political parties. Even where self-censorship is the aim of a technological system (as with the Chinese Social Credit scheme, which is designed to create citizen behaviour that aligns with governmental priorities (Creemers, 2016)), it can be argued that it is not realistic for users to remain constantly in a state of struggle against an all-encompassing system of surveillance. Instead, evidence shows that the increasing necessity of data technologies in everyday life causes people to resign themselves to this distributed visibility rather than engaging with it politically (Turow et al., 2015).

Until now, within the global North freedoms and needs with regard to data technologies have been approached through a fundamental rights framework that includes data protection, framings of informational privacy and the right to free speech and communication. However, this framing presents two problems when applied in relation to the global data market. First, the liberal individual framing of Human Rights requires that abuses are clear and visible so that those injured can respond, and second, it assumes that redress will be sought on the individual level. This is rendered problematic by the invisible and many-to-many nature of 'seeing' through data technologies, but also by the fact that many of the negative impacts of data occur on the group as much as the individual level (Taylor et al., 2017).

Instead of applying a fundamental rights framework whose application demands identifiable violations, this new situation requires a more multifaceted approach that can address the breadth of actors and possibilities inherent in contemporary data collection and use. By identifying the new ways in which power is inscribed in large-scale digital data, we can better debate what we want and do not want from the information we emit about ourselves. The next section will explore two examples from the public–private interface of datafication that illustrate the ways in which that interface may be a locus of structural discrimination (embedded in institutions, rules and practices) that is also intersectional (multiplying disadvantage to people due to intersecting aspects of their identity).

Identifying data injustices

I will explore the problem of data-driven discrimination using two illustrative cases, both of which demonstrate that a specific articulation of social justice is now required with regard to contemporary data technologies. The first case is that of India's biometric population database, known as Aadhaar. The database is the world's largest with over a billion records and was launched in 2009 with the stated aim of combating welfare fraud by allowing those below the poverty line to prove their identity with a fingerprint or iris scan when collecting entitlements. However, the design of the technologies that enable inclusion in the system – iris and fingerprint scanners and the networks, wired and human, that translate inputs of data into outputs of confirmed identities – in fact ensures that the poorest are the worst served by Aadhaar.

The system's design does not acknowledge the materiality of poverty, being unable to 'authenticate those who work with stone, cement, limestone and those over the age of 60' (Yadav, 2016) since they often have no fingerprints due to hard labour, or usable iris scans due to malnutrition. It also misses the day-to-day precarity of poor people's existence by only allowing each family's single registered claimant to draw rations, so that if that claimant is sick, working or otherwise unable to come to the ration provider, the family cannot access its allocation (Priya and Priya, 2016). Moreover, the backup authentication system operates by sending a password to the registrant's mobile phone, thus excluding the poor who cannot afford a phone, or anyone who has not written down the number they had when they enrolled (Yadav, 2016). It also increases the bureaucratic burden of poverty, since despite compulsory participation there is no way for people to correct entries in the database on the local level. There is no independent oversight in terms of addressing technological faults: the redress system refers people back to the Unique Identification Authority of India, the agency that runs Aadhaar, but there is no legal obligation for the authority to provide a solution to authentication problems, leaving them instead to individual citizens and local ration shops to resolve (Thikkavarapu, 2016). Despite its

unresponsiveness to registrants, the database does, however, make it possible for the ultra-poor to be transformed into consumers: its chairman has said that he envisages it as having strong potential for direct marketing to registrants (Nilekani, 2013) and plans are underway to partner with Google so the firm can reach and profile its 'next billion users' (Aulakh and Surabhi Agarwal, 2016).

The problem of Aadhaar's uneven burden on the poor was illustrated by the government's 2016 demonetisation, which threw the cash-based elements of India's economy into chaos, and thus also the lives of the poor and marginalised. The demonetisation put demands on automated payment systems in ways that were discriminatory against the poor, since they had the least access to mobile phones, formal saving and banking systems, and applications that could help tide them over the cash crisis that ensued – and suffered the highest cost if Aadhaar-related technologies failed to identify them correctly (Masiero, 2017).

Aadhaar is what Johnson (2014) in his work on information justice terms a 'disciplinary system'. It raises several issues to do with justice that are specific to its use of data technologies, specifically the way it records, stores and processes registrants' data. First, at the point of collection and processing claims the system forces registrants to confirm to a 'standard of normalcy' (Johnson, 2014) by having legible fingerprints and irises, by possessing mobile phones, by having a stable family life where the same registrant can collect rations from week to week, among other standards. These standards point to a middle-class standard for normality rather than the precarity and unpredictability of the lives of the poor. Second, it raises problems of distributive fairness. Although the claim is made that Aadhaar furthers distributive justice by reducing corruption in welfare transactions and giving the poor access to previously inaccessible services and representation, in fact it offers radically different possibilities depending on one's resources and socio-economic status. Third, the system amplifies inequality: for richer citizens, it is a way to ease one's passage through the world. One can acquire a phone or a utility account, prove one's identity in everyday transactions and simplify dealings with the bureaucracy. For poorer citizens, often lower caste and/or female, it is a way of formalising precarity. For those whose bodies the system cannot process, or for those whose identity is misread, there is no apparent path back to administrative legibility. Finally, it does not allow for fair redress of abuse or grievances. The complaints procedure is not designed for emergencies: instead of access to a local official, problems must be directed by phone or email to a processing centre with 'no timelines, no designated grievance redress officers, no written dated acknowledgement receipts, no compensation for the complainants and no penalties for erring officials' (Sabhikhi, 2016). Although it is only the poor who have no choice about whether to use the system, it is aligned with the bodies and lifestyles of the middle and upper classes. Meanwhile, there are reports of growing malnourishment amongst families excluded by the database (Priya and Priya, 2016).

A second example comes from a system that at the time of writing was still at the conceptualisation stage, but which demonstrates how algorithmic uncertainty (Kwan, 2016) – the gap between virtual spatial information and physical ground truth – can translate into embedded injustice. A recent proposal by a commercial firm consulting for the EU Space Agency aimed to monitor the movements of migrants moving towards the EU's southern borders.[1] Using machine learning performed on satellite images showing groups as they prepare to board boats to cross the Mediterranean, social media output and local online reporting, the firm proposed to track migrants and predict their origin and direction of movement. The consultants planned originally to use migrants' mobile phone traces, but were dissuaded by the difficulty of obtaining ongoing datasets, which are tightly guarded by mobile operators due to privacy concerns (Taylor, 2016c). The objective of the proposed project was described as being to enable the visualisation of migrants heading towards Europe by identifying small groups on individual beaches or hillsides and predicting who would cross which border and when. The predictions, sold on to border enforcement and migration authorities, would then potentially allow those authorities to use algorithmic sorting to identify 'undesirable' migrants and control the numbers able to make asylum claims by putting in place measures to prevent them reaching European soil.

This proposal was problematic for several reasons. First, the machine learning involved would allow the project to categorise migrants based on behaviour and characteristics recorded remotely, and second, the results of that analysis would then be channelled to institutions who decide whether to allow those migrants to enter or not. The first aim is risky because it involves using other attributes as proxies for the targeted behaviour or characteristics – for example, assuming that people congregating on a particular beach at a particular time, or who have been posting certain keywords or terms on social media, may be planning to migrate and claim asylum in a particular place. In turn, these proxies for place of origin and direction of travel are designed to be used to predict the likelihood that a group of migrants will have a valid asylum claim.

In reality, however, it is possible to be an undocumented migrant from anywhere on earth and have a valid claim to asylum. This is because anyone can be

at risk even in a 'safe' environment. For example, during the US government's attempt to impose a travel ban on Muslims during 2016, the Canadian government considered whether to declassify even the US as a safe country for refugees (Kassam, 2017). Citizens of democratic countries that are not at war may be at risk due to their sexuality, their gender, their religious or ethnic group, their political affiliation or any number of other characteristics. Conditions of risk to individuals are contingent, shifting and nuanced far beyond what can be predicted by capturing and weighing proxies for place of origin and migration behaviour. Such a model, based on the long-discredited idea of 'acting suspiciously', will almost inevitably eventually be used to produce a yes/no answer that, used on the group level, determines life or death for individual migrants.

A system such as this serves to demonstrate how the possibilities of control through data technologies coevolve with the possibilities of care (Lyon, 2007). Refugees use similar technology to guide their path into Europe, but in ways that protect their identities and give them some degree of control over their trajectory. They share their phone GPS details with relatives and volunteer helpers, use Google Maps to find their way and social media to decide how to make their journey (BBC News, 2015; Ram, 2015). When used by individuals and groups on the ground, the same satellite-based GPS and mapping technologies that can be used by border agencies to control and erase migration also help migrants to preserve autonomy, human security and their right to flee danger.

There is a strong incentive for commercial analysts, in this case, to stress the accuracy of their automated predictions, since this raises their value for public-sector buyers. A 'good usable system', as Bowker and Star (1999: 33) have observed, becomes so convenient that it disappears and only its answers can be seen. In the context of large-scale remote surveillance for policy purposes, the likelihood that those answers will be tested for accuracy is diminished. O'Neil (2016) has warned that algorithmic models must be constantly recalibrated using feedback from the events they are supposed to predict. Yet it is hard to see how a model that uses remote surveillance to predict the aims and origins of undocumented migrants will also incorporate correct data on those migrants' actual outcomes.

The likely inaccuracy of such a system notwithstanding, this leads us to the larger problem: that rights that are supposed to be fundamental – including the rights to privacy and autonomy, to effective remedy for harms, and many others relating to the uses of data in a migration-prediction system – are in fact not treated as fundamental because they do not extend across borders. The problem of migrant rights goes far beyond data, but data systems underpin the way migrants can be included and excluded, and therefore their ability to claim their rights (Broeders, 2009). Data systems that are transnational in nature demand rights and redress mechanisms that are similarly transnational, yet this is currently an impossibility in the case of remote sensing, remote analytics and remote decision making. Someone who is remotely surveilled will not know the basis on which they are being categorised, and in any case those who are territorially excluded as a result will not be able to appeal the decision. Yet if the functioning of national law and sovereignty mean that a person has certain fundamental rights when standing on one side of the EU's border but not on the other, this is deeply problematic in ways that cannot be solved by claims about the necessity of sovereignty (Brock, 2009). Data justice joins a class of complex multidimensional problems such as climate justice, terrorism and poverty that have been classed as 'super-wicked' (Levin et al., 2012), and which it is necessary to address in a systemic way in order to deal with their interdependencies.

Data justice across domains and systems

There are currently (at least) three main approaches to conceptualising data justice: one addressing the ways in which data used for governance can support power asymmetries (Johnson, 2014), another focusing on the ways in which data technologies can provide greater distributive justice through making the poor visible (Heeks and Renken, 2016) and another that is interested in how practices of dataveillance can impact on the work of social justice organisations (Dencik et al., 2016). Although these different strands of research apparently point in different directions, I will argue that there is value in bringing them together.

Johnson (2014) connects data justice primarily to open data. He writes of the need for 'subsuming the question of open data within a larger question of information justice', but goes on to offer conclusions that have relevance for data as a tool of governance (and governmentality) more broadly. He advocates for establishing a concept of 'information justice' that can counteract the ways in which administrative data inevitably embeds social privilege and creates an unequal set of opportunities due to the differential capabilities of citizen versus commercial users. He argues that data systems tend to have a disciplinary function because the way data are collected and structured constitutes a form of normative coercion (one example of this is the problem encountered by transgender people seeking to change their birth registration as cited above). As a way to address the problem, Johnson (2016: 29)

advocates for 'mak[ing] politics explicit' with regard to data technologies, through collaborative research involving philosophers of technology, information scientists and social scientists.

Second, Heeks and Renken (2016) provide a rich analysis of the possible framings of data justice from the perspective and with the priorities of the international development sector, taking the question of information justice to an explicitly global level and asking how it should be purposed when applied to questions of human development. The paper begins from the notion that the Sustainable Development Goals forefront both 'data' and 'justice', and that therefore the development field must engage with them as intersecting concepts for the first time. The authors argue for a structural approach that does not limit itself to the functions of data in the sector, but instead is framed with reference to 'wider codes of social and political justice' (p. 5). They use the Universal Declaration of Human Rights (UN General Assembly, 1948) to argue that the rights to data ownership, access and representation are fundamental to fairness and justice. The paper argues for a networked perspective, seeing data systems as connectors on the local and global levels where competing interests are inevitably at play.

Finally, Dencik et al. (2016) identify a need to conceptualise data justice due to the way that surveillance capitalism constrains citizenship and activism. They therefore argue for the introduction of 'data justice' terminology to describe resistance to government surveillance based on principles of social justice. Their idea of data justice pays attention to the way in which choices about data systems, contractors and targets inscribe particular kinds of power and interest. In their framing, data justice is a concept that can help create collaboration between anti-surveillance and social justice activism, driving the first to articulate broader concerns of rights and freedoms and the latter to engage with the technical dimensions of surveillance and resistance. Their framing focuses specifically on social activism, but connects both to Johnson's call to explore the politics of datafication and to Heeks and Renken's attention to the political economy of datafication: 'Referring to ''data justice'' recognises the political economy of the system that underpins the possibilities for extensive surveillance, whilst drawing attention to the political agenda that is driving its implementation' (Dencik et al., 2016: 10).

These three contrasting interpretations of the idea of data justice are linked through their focus on politics and power, and by their formulation of social justice. The differences between their conceptualisations, however, are useful because they raise some essential questions.

First, how can a conceptualisation of data justice on the global scale call on important fundamentals such as rights, justice and fairness without becoming relativistic? Heeks and Renken (2016: 7) note that each region or country will judge what is just according to its own conceptualisation based on its own tradition and history, and that they therefore seek 'to move right away from interpretive, bottom-up notions' to wider codes such as the UDHR.

How is it possible, though, to formulate principles of data justice without allowing them to be shaped by the global community of data producers? A vision of data justice that takes power and politics into account must necessarily also be rooted in local experience. If countries have differing aims with regard to the potential of digital data, and different ideas of what constitutes its misuse, how should they contribute to the framing of what is just? For example, the argument has been made the subjects of development effectively have a duty of visibility to the authorities working to combat poverty (Taylor, 2016c). Robert Kirkpatrick of the UN's Global Pulse data science initiative has said of developing-country citizens that 'privacy is your right. So is access to food, water, humanitarian response. The challenge is that we see a lot of regulatory frameworks which don't have the right litmus test'.[2] His statement implies that development agencies have a claim to people's data on a utilitarian basis, and that opting out should not be an option because it will impact on the rights of the collective.

Up to this point, I have used the words 'rights' and 'freedoms' to denote the concepts that seem essential as metrics for the just use of data technologies. The Global Pulse example, however, indicates the need for a relational approach (as opposed to a relativistic one) that can integrate rights and needs into a single framework rather than demand a purely utilitarian choice between them. In fact the language of rights may not be the right tool with which to seek to define a global framing of justice. Brock (2009) argues in her cosmopolitan account of global justice that asking what people's needs are, rather than what rights they may claim, makes it possible to think across cultural and regional framings of justice. The liberal individual notion of rights is not adopted by many societies which nevertheless have a strong philosophical and legal tradition of justice and fairness (Panikkar, 1982). Many states and regions, for example, take a perspective on rights that frames them as inseparable from corresponding duties, and also addresses the individual as part of the larger collective (Sen, 2005).

Datafication is frequently a territory for both the formalisation and the negotiation of rights. For example, in the case of many African states Makulilo (2016) posits that with the arrival of the digital era, the

growing data economy is placing the notion of *ubuntu* (humanity towards the collective) into a complex interaction with that of privacy. Arguing that it is possible to find certain freedoms identified as important across different societies in the region regardless of their formalisation as human rights, Makulilo quotes the Nigerian scholar Nwauche, who states that despite the lack of a formal right to it, 'privacy is important in Nigeria because there are human beings'. On this basis, it may be more useful to think in terms of basic needs with regard to data that may be formalised differently in different places.

Also mitigating against the use of a framework based on individual rights to shape data justice is the fact that data injustice increasingly tends to occur on the collective level. New data technologies tend to sort, profile and inform action based on group rather than individual characteristics and behaviour (Taylor et al., 2017), so that in order to operationalise any concept of data justice, it is inevitably going to be necessary to look beyond the individual level. Some legal systems have already formalised this relationship between surveillance and the collective: Mexican law, for example, envisages this collective level with regard to informational privacy by including the family in the sphere of the individual with regard to data protection (Tribunales Colegiados del Circuito, 2015). Sen (2005), whose work is central to Heeks and Renken's exploration of data justice, balances the individual and collective when he notes that social justice cannot occur in a vacuum but requires collective action to be realised. The structural approach is key here: if we consider the necessity for the formal establishment of what is fair (process freedoms) and for the need for agency (opportunity freedoms) (Sen 2005), data justice seems to fit best with a broader capabilities perspective that can encompass variation in how fairness should be determined and whether justice can be realised.

A second issue that arises from the three framings quoted above is the issue of what data justice should aim to achieve. The three visions focus in very different directions: Johnson asks how database design can better incorporate anti-discrimination principles; Heeks and Renken focus mainly on the question of how data should be distributed in order to achieve fairer access, participation and representation. Dencik et al., meanwhile, are concerned with the conditions under which data should *not* be distributed (via surveillance), in order to protect the work of activists working towards social justice. A further, contrasting perspective can be found in the work of Mann (2016), who argues that it is also important who gets to process the economic benefits of the data economy, and that to promote social justice in relation to digital data, we should seek to embed principles of justice in the way data markets are structured.

Are these perspectives contrasting or incompatible? Do we need separate frameworks for developing better defences against discrimination, exercising the right to be counted and resisting surveillance, or is it more useful to find an overarching argument about data justice that can incorporate these principles and, by doing so, contribute more than addressing them separately? In the next section I will argue for such an overarching framework in order to reconcile these aims, each of which promotes a different but essential freedom with regard to data.

A proposed framework for data justice

A framework that could reconcile the differing perspectives discussed above would have to do several things. First, it would have to take into account the novelty and complexity of the ways in which (big) data systems can discriminate, discipline and control, as demonstrated in the examples of Aadhaar and the migration monitoring system. Second, it would have to offer a framing which could take into account both the positive and negative potential of the new data technologies – their ability to facilitate what Nussbaum and Sen (1993) term 'human flourishing', which forms the overall aim of human development – and also their potential to hinder it. Finally, it would have to do this using principles that were useful across social contexts, and thus remedy the developing double standard with regard to privacy and the value of visibility in lower versus higher income countries, as illustrated by Global Pulse's utilitarian balancing of privacy with other needs.

A framework is necessary, then, that can take into account the need to be represented but also the possibility of the need to opt out of data collection or processing, the need to preserve one's autonomy with regard to data-producing technologies and the need to be protected from and to challenge data-driven discrimination. This suggests an approach based on three pillars: visibility, digital (dis)engagement and countering data-driven discrimination (see Figure 1), but which does more than set out what rights are necessary. It must also provide for a methodological engagement with the political economy of data, in order to determine not only *what*, but *who* is important and *how* they relate to the desired outcomes.

As Figure 1 shows, the elements of data justice hypothesised here are purposely broader than available international frameworks such as the Fair Information Practice Principles which form the basis for informational rights in many OECD countries, or the right to privacy as set out in the various human rights instruments. These frameworks are valuable and often effective, but they aim at the problem on a practical rather

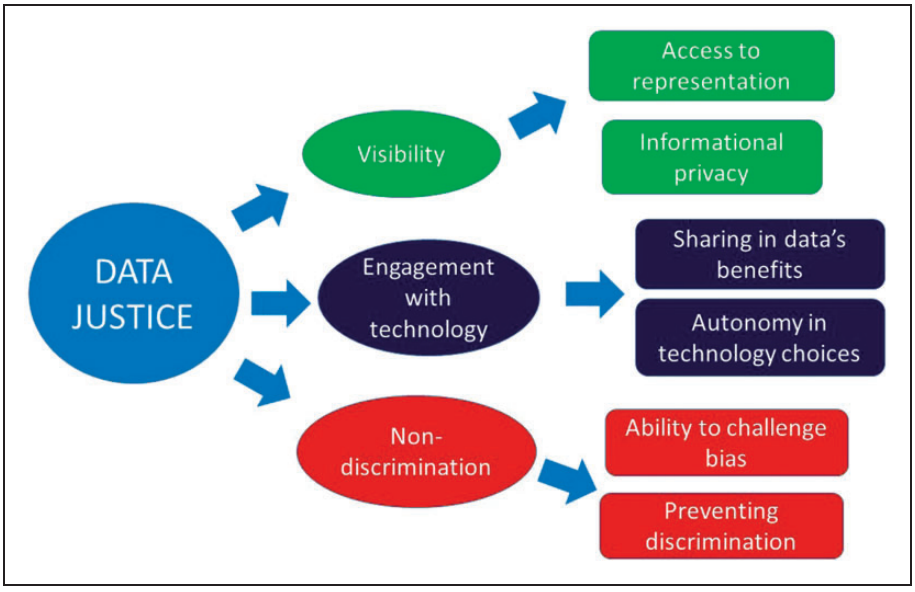

Figure 1. Three pillars of data justice.

than conceptual level. Rather than staking out the boundaries of what should or should not be done with data, the elements of data justice proposed here represent a way to think about data at a level that goes beyond particular domains and applications, and instead to address data technologies primarily as they relate to human needs.

The first pillar, visibility, deals both with privacy and representation. Here common threads of reasoning can be drawn from the fields of international development studies, human geography and legal scholarship. A more detailed framing of the needs for both visibility and informational privacy should take into account the work being done on privacy at the social margins (Arora, 2016; Gilliom, 2001; Jayaram, 2014), the risks to group privacy through collective profiling (Taylor, 2016b; Floridi, 2014; Raymond, 2016) and the extent to which data may be considered a public good (Taylor, 2016d).

Engagement with technology is the second pillar of this putative conceptual framework. Although ICT-for-Development – the promotion of engagement with digital technologies in LMICs – has clearly established links between fostering human development and providing access to ICTs (Heeks, 2010; Unwin, 2009), this field is adjusting, like others, to the evolution of new data-producing technologies and analytics and is now starting to critically address the freedom *not* to use particular technologies, and in particular not to become part of commercial databases as a by-product of development interventions (Taylor & Broeders, 2015; Gagliardone, 2014). The freedom to control the terms of one's engagement with data markets is an essential component of any data justice framework because it underpins the power to understand and determine one's own visibility. Arguments for the importance of people's autonomy with regard to technology can be found in postcolonial theory, since the way in which data is processed and analysed within national and global data markets positions individuals as subalterns (Spivak, 1988) in relation to those who process their data. They are unable to define for themselves how their data are used, to whom they are resold or the kinds of profiles and interventions those data can enable. On this basis Mann (2016) argues that addressing *data for economic development*, rather than just data for development per se, focuses attention on the potential benefits to low-income people of collecting and analysing data independently of large technology firms, and on how data's returns can be captured and processed at the local level.

The third pillar within this proposed framework is nondiscrimination. It is composed of two dimensions: the power to identify and challenge bias in data use, and the freedom not to be discriminated against. People's ability to identify and challenge bias in data-driven decision making is expected to diminish as the complexity of data's production and processing increase (Kroll et al., 2016). As neural networks and deep learning become more commonplace, the ability of even system designers themselves to understand how bias may be embedded in data processing is diminishing. This implies that methods have to be devised that can allow for the governance of algorithmic processes and decision making, and that the responsibility for challenging discrimination on the part of individuals will need to be accompanied by the ability to identify and create penalties for it on the part of government (Kroll et al., 2016).

Part of the contribution of the proposed conceptual framework for data justice is to help frame the necessary questions as well as to point the way to answers. As in the periodic table, any mapping of relationships and positions also maps out missing elements. For example, the *freedom not to engage with the data market* or *not to be represented in commercial databases* has not yet been adequately theorised: even privacy studies assume that such engagement is inevitable. Yet there are historical and contemporary indications that such freedoms are necessary: From the 19th-century Luddite protestors who developed a new politics of workers' rights and resistance in the context of new industrial technologies (Binfield, 2015) to the anti-census activists of the 1970s and 1980s in the Netherlands and Germany (Hannah, 2010; Holvast, 2013), there has been a debate about where technology that can count and monitor fits within the social contract and how much visibility citizens owe the state. This debate underlies both Heeks and Renken's and Johnson's formulations: the right to be seen and represented is central to data justice, but so is the right to withdraw from the database, whether it is held by the state, commercial firms or both as in the case of Aadhaar.

Perhaps the central question raised by the concept of data justice set out here is how to balance and integrate the need to be seen and represented appropriately with the needs for autonomy and integrity. What are the implications of letting people opt out of data collection? Should people, for example, be able to opt out of commercial databases if those databases are likely to be used by the state to supplement or replace administrative or survey data? What are good governance principles for the use of big data in a democratic context, and who should be responsible for determining them? Census-taking is both one of the most invasive moments in the relationship between individual and state, and one of the most important rights of a citizen in a democratic society. If state population data is soon to be at least partly composed of commercially collected data (Keeter, 2012) and updated in real time, and those data can tell the government not only conventional facts about the population but instead almost everything, where does legitimate observation end and illegitimate surveillance begin?

An ecosystemic approach based on capabilities

As Heeks and Renken suggest, Sen (1999) and Nussbaum's (2006) capabilities and freedoms-based approach offer one avenue to integrating the principles of data justice into an operationalisable framework, as does Kleine's (2011) Choice Framework for conceptualising the opportunities provided by technology in a development context. The Capabilities Approach encompasses both what Sen terms *opportunity freedoms*, described by Alkire (2011) as one's real opportunity to achieve the functionings one values, and what he terms *process freedoms*, which denote agency – one's ability to act on behalf of what matters (Alkire 2011). In line with the focus of data justice on preventing marginalisation and promoting a socially just model for handling data, the approach begins not from a consideration of the average person, but asks instead what kind of organising principles for justice can address the marginalised and vulnerable to the same extent as everyone else (see Figure 2).

The diagram demonstrates how data justice can fit within the Capabilities Approach as an overarching conceptual framework within which research and debate can identify what freedoms people value with regard to data technologies, and how to realise them. It is positioned within the structure of opportunity freedoms and process freedoms that determine what people's *functionings* ('doings' or 'beings') can be with regard to data technologies. In turn, these functionings can be translated via social *conversion factors* such as political, legal and educational support into *capabilities* such as participation in data value chains, access to data affecting oneself (e.g., through laws on freedom of information or via sectoral data access facilities) and inclusion in decision making about what technologies are used in particular contexts.

If we follow Sen's advice to engage with the domain of public reasoning to determine what people want from data technologies, this also leads us to consider his overall argument for a Capabilities framing: that it helps people to decide what functionings they value, and what capabilities they wish to prioritise. The task of a conceptual framework for data justice, then, becomes to build on this approach to understand what constitute the common principles that might help operationalise it. Such principles are necessary with regard to the global nature of the data market, since national legislation has difficulty targeting processes that take place transnationally, as is the case with, for example, data brokers and large online service providers such as Google and Facebook.

The task of building out and thinking about operationalising this kind of data justice framework therefore requires a different theoretical and methodological toolkit from, for instance, research on informational privacy. This fits with a current trend where scholars worldwide are calling for change in terms of the way data's social impacts are researched (Cohen, 2012; Dalton et al., 2016; Floridi, 2016; Kleine, 2010; Schwartz and Solove, 2011). Cohen (2012), in particular, has argued for a socially situated and

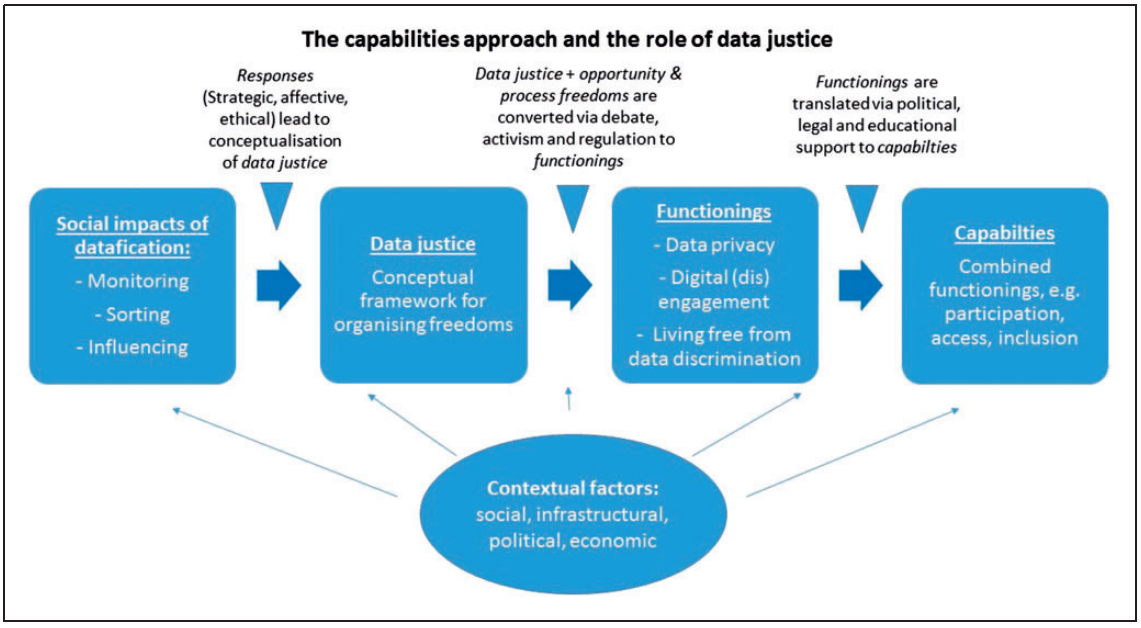

Figure 2. A capabilities approach to data justice.

interdisciplinary analysis of information, power and privacy that can create new organising notions for human flourishing relevant to the rise of the 'networked self'. Such research would need to take a global ecosystemic approach that could look across borders. It would offer the possibility of bridging among different levels of engagement with technology and different concepts of technology-related development, and between different moral and philosophical systems. One important tool in this process would be the emerging field of Critical Data Studies, and that of digital geographies more broadly, which have shown that the knowledge necessary to establish a more socially just approach to the use of digital data already exists, but that it tends not to be incorporated into policy, law or practice at the level necessary to be usable (Dalton et al., 2016). Connecting this knowledge to policy and law, and particularly to the transnational political responsibilities of online service providers (Taylor et al., 2017) would inevitably be part of the work of operationalising data justice, just as it is with social justice movements more broadly.

The questions raised by the framing of data justice presented here operate both at the highest level – where the social contract is shaped and negotiated – and at the most basic, in the practices of everyday digital life. These, as Heeks and Renken point out, will differ across societies and regions. Therefore, the main challenge in building out this conceptualisation will be in finding how these overarching principles can gain traction in different contexts: some countries or groups will identify benefits of surveillance while others will strongly react against it as oppressive. Some will assert that private sector innovation plays a central role in realising the benefits of data science while others will claim that making the public sector more responsible for controlling data will achieve more just results. The conceptual core would need to be translated and negotiated across contexts just as its components already have been (e.g., data protection or research ethics). Ideas such as justice, equality and non-discrimination inform regimes as varied as taxation and market regulation, so that data justice would need to operate as another branch of these core governance principles. Rather than being centralised however (the 'centre' inevitably being a high-income, high-technology location), these translations and negotiations would instead have to occur on a distributed basis within what Sen (2005) has termed 'the domain of public reasoning'. Under such a premise, each legal and social system would work out for itself how the principles of data justice applied. This is important because the principles set out here are at odds, in one way or another, with every established regime of data governance on earth (e.g. the right not to be recorded in databases), and will meet different challenges depending on the location of the discussion.

Conclusion

The conceptualisation of data justice presented above poses a challenge to most existing frameworks for governing data. It does so because it incorporates the assumption that any framing that does not incorporate both the beneficial and negative aspects of data technologies cannot gain traction in the domain of public

reasoning. The frameworks we currently use either emphasises risk and harm, or argue for making data and the power to analyse them as broadly accessible as possible. The task of reconciling these perspectives is politically and theoretically huge. A framework that aims to reconcile datafied visibility with invisibility and technological engagement with nonengagement will challenge many accepted norms, notably in the areas of promoting innovation and economic development and the established right of the state to count and intervene upon its citizens. The principles set out here are not obstacles to innovation, nor should they constitute a hindrance to democratic processes of government. Nevertheless, they pose difficult questions that require significant reconciliation of different values. It is important to pose these questions because they aim at the changing interfaces between the individual and the state, between the commercial and public sectors and between science and the public, and they mark out the uncomfortable territory where friction is taking place around privacy, responsibility and accountability.

These places of friction deserve our attention. They are the places where we are negotiating both the evolution of governance and of how we wish to live alongside each other in knowledge societies. Change should not take place at these intersections without our noticing, nor should we brush it off as inevitable. Innovation and evolution in technology are constant and desirable, but the ways in which technologies are used to monitor and govern us are negotiable. We should be able to determine our interactions with technology by debating and, if necessary, resisting and proposing different paths. If we cannot imagine how to regain the kind of privacy we would like, how to allow people to opt out of being surveilled through their data – or even of producing those data in the first place – we may have to reinvent as well as renegotiate.

This may also involve making different demands of authorities – whether commercial or governmental – with regard to the governance of, and through, data technologies. Operationalising the framework proposed here would entail a shift from making individuals responsible for understanding the data market to making national and international authorities accountable for data governance. It would also demand that we distinguish between *responsible* data use – the current buzzword in the fields of data governance and innovation policy – and *accountable* data use, something much more difficult to achieve because it demands structural change rather than allowing our guardians to guard themselves.

The various framings of data justice proposed since the advent of big data indicate that around the world, scholars and policymakers are attempting to reconcile principles of social justice with the reality of datafication. Their contributions range from the grand scale of the Onlife Manifesto to the specificity of Greenfield's *Against the Smart City* (2013). The next challenge is to integrate these worldwide perspectives and principles into a broader vision that can address the globalisation of data technologies and its impacts. The framework set out here is one response to the challenge of making sense of life in datafied societies. It aims to offer a roadmap towards further analysis, the specification of particular aims and objectives, and eventually operationalisation within multiple and differing national and international contexts.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The research was conducted at the Oxford Internet Institute and the University of Amsterdam, at the latter under a Marie Curie postdoctoral fellowship (624583 D4D).

Notes

1. This example is based on personal communication during 2016 between the author and those planning the project.
2. Robert Kirkpatrick, UN Global Pulse, interviewed 18 August 2014.

References

Alkire S (2011) The capability approach and human development. University of Oxford. Available at: http://www.ophi.org.uk/wp-content/uploads/OPHI-HDCA-SS11-Intro-to-the-Capability-Approach-SA.pdf (accessed 23 June 2017).

Arora P (2016) Bottom of the data pyramid: Big data and the global south. *International Journal of Communication* 10: 19.

Aulakh G and Surabhi Agarwal N (2016) Google in talks with government to partner for Aadhaar, UPI. *The Economic Times*. Available at: http://economictimes.indiatimes.com/opinion/interviews/google-in-talks-with-government-to-partner-for-aadhaar-upi-caesar-sengupta-vice-president-next-billion-users-at-google/articleshow/54556320.cms (accessed 28 September 2016).

BBC News (2015) Migrant crisis: ''We would be lost without Google maps.'' Available at: https://www.youtube.com/watch?v=Zcr-GWv3Qbs (accessed 22 June 2017).

Binfield K (2015) *Introduction*. Writings of the Luddites. Baltimore, MD: Johns Hopkins University Press.

Bowker GC and Star SL (1999) *Sorting Things Out: Classification and Its Consequences*. Cambridge: MIT Press.

Brock G (2009) *Global Justice: A Cosmopolitan Account*. Oxford: Oxford University Press.

Broeders D (2009) *Breaking Down Anonymity: Digital Surveillance of Irregular Migrants in Germany and the Netherlands*. Amsterdam: Amsterdam University Press.

Broeders D and Taylor L (2017) *Does great power come with great responsibility? The need to talk about corporate political responsibility*. The Responsibilities of Online Service Providers. New York: Springer, pp. 315–323.

Cho S, Crenshaw KW and McCall L (2013) Toward a field of intersectionality studies: Theory, applications, and praxis. *Signs: Journal of Women in Culture and Society* 38(4): 785–810.

Cohen JE (2012) *Configuring the Networked Self: Law, Code, and the Play of Everyday Practice*. New Haven, CT: Yale University Press.

Creemers R (2016) What could China's "social credit system" mean for its citizens? *Foreign Policy*. Available at: http://foreignpolicy.com/2016/08/15/what-could-chinas-social-credit-system-mean-for-its-citizens/ (accessed 26 June 2017).

Dalton CM, Taylor L and Thatcher J (2016) Critical data studies: A dialog on data and space. *Big Data & Society* 1–9. doi: 10.1177/2053951716648346.

Dencik L, Hintz A and Cable J (2016) Towards data justice? The ambiguity of anti-surveillance resistance in political activism. *Big Data & Society* 3(2): 1–12.

Eubanks V (2014) Want to predict the future of surveillance? Ask poor communities. *The American Prospect*, pp. 1–4. Available at: http://prospect.org/article/want-predict-future-surveillance-ask-poor-communities#.VXbsO_Oh2k8.twitter (accessed 9 October 2017).

European Commission (2016) *Digital Agenda for Europe*. Available at: https://europa.eu/european-union/file/1497/download_en?token=KzfSz-CR.

Floridi L (2014) Open data, data protection, and group privacy. *Philosophy and Technology* 27: 1–3.

Floridi L (2016) On human dignity as a foundation for the right to privacy. *Philosophy and Technology* 1(6). Available at: https://www.youtube.com/watch?v=CD5zfBcAHms&feature=youtu.be (accessed 9 October 2017).

Foucault M (1977) *Discipline & punish – Panopticism*. Discipline and Punish: The Birth of the Prison. New York: Vintage, pp. 195–228.

Gagliardone I (2014) "A country in order": Technopolitics, nation building, and the development of ICT in Ethiopia. *Information Technologies and International Development* 10(1): 3–19.

Gilliom J (2001) *Overseers of the Poor: Surveillance, Resistance, and the Limits of Privacy*. Chicago, IL: University of Chicago Press.

Greenfield A (2013) *Against the Smart City: A Pamphlet*. Do Projects.

Hannah M (2010) *Dark Territory in the Information Age: Learning from the West German Census Controversies of the 1980s*. Burlington, NJ: Ashgate.

Heeks R (2010) Development 2.0: The IT-enabled transformation of international development. *Communications of the ACM2* 53(4): 22–24.

Heeks R and Renken J (2016) *Data Justice for Development: What Would It Mean?* Manchester. Available at: https://www.gdi.manchester.ac.uk/research/publications/other-working-papers/di/di-wp63/ (accessed 9 October 2017).

Holvast J (2013) *De Volkstelling van 1971*. Amsterdam: Uitgeverij Paris.

ITU (2015) Key ICT indicators for developed and developing countries and the world (totals and penetration rates). Available at: https://www.itu.int/en/ITU-D/Statistics/.../2014/ITU_Key_2005-2014_ICT_data.xls (accessed 9 October 2017).

Jayaram M (2014) India's big brother project. *Boston Review*. Available at: http://www.bostonreview.net/world/malavika-jayaram-india-unique-identification-biometrics (accessed 1 January 2016).

Jiwani Y (2015) Violating in/visibilities: Honor killings and interlocking surveillance(s). In: Dubrovsky RE and Magnet SA (eds) *Feminist Surveillance Studies*. Durham, NC and London: Duke University Press, pp. 79–92.

Johnson J (2014) From open data to information justice. *Ethics and Information Technology* 16(4): 263–274.

Johnson J (2016) The question of information justice. *Communications of the ACM* 59(3): 27–29.

Kang LHY (2015) Surveillance and the work of anti-trafficking: from compulsory examination to international coordination. In: Dubrofsky RE and Magnet SA (eds) *Feminist Surveillance Studies*. Durham, NC and London: Duke University Press, pp. 39–57.

Kassam A (2017) Refugees crossing into Canada from US on foot despite freezing temperatures. *The Guardian*. Available at: https://www.theguardian.com/world/2017/feb/07/us-refugees-canada-border-trump-travel-ban (accessed 26 June 2017).

Keeter S (2012) Survey research, its new frontiers, and democracy. *Public Opinion Quarterly* 76(3): 600–608.

Kitchin R (2016) The ethics of smart cities and urban science. *Philosophical Transactions of the Royal Society A* 374(2083): 1–15.

Kleine D (2010) ICT4WHAT? – Using the choice framework to operationalise the capability approach to development. *Journal of International Development* 22(5): 674–692.

Kleine D (2011) The capability approach and the "medium of choice": Steps towards conceptualising information and communication technologies for development. *Ethics and Information Technology* 13(2): 119–130.

Kroll JA, et al. (2016) Accountable algorithms. *University of Pennsylvania Law Review* 165(633): 633–705.

Kwan M (2016) Algorithmic geographies: Big data, algorithmic uncertainty, and the production of geographic knowledge. *Annals of the American Association of Geographers* 106(2): 274–282.

Lemke T (2001) "The birth of bio-politics": Michel Foucault's lecture at the Collège de France on neo-liberal governmentality. *Economy and Society* 30(2): 190–207.

Levin K, et al. (2012) Overcoming the tragedy of super wicked problems: Constraining our future selves to ameliorate global climate change. *Policy Sciences* 45(2): 123–152.

Lyon D (2007) *Surveillance Studies: An Overview*. Cambridge: Polity Press.

Makulilo AB (2016) "A person is a person through other persons" – A critical analysis of privacy and culture in Africa, *Beijing Law Review* 7: 192–204.

Mann L (2016) Corporations left to other peoples' devices: A political economy perspective on the big data revolution in

development. *Development and Change*. Epub ahead of print. doi: 10.1111/dech.12347.

Masiero S (2016) Digital governance and the reconstruction of the Indian anti-poverty system. *Oxford Development Studies* 818: 1–16.

Masiero S (2017) Will Aadhaar help the poor become cashless? *LiveMint*. Available at: http://www.livemint.com/Opinion/Mj3KgqCK1cZ2hbZYwbvE9H/Will-Aadhaar-help-the-poor-become-cashless.html (accessed 15 February 2017).

Moore LJ and Currah P (2015) Legally sexed: Birth certificates and transgender citizens. In: Dubrovsky RE and Magnet SE (eds) *Feminist Surveillance Studies*. Durham, NC and London, pp. 58–78.

Nilekani N (2013) Technology to leapfrog development: The Aadhaar experience. Available at: http://www.cgdev.org/sites/default/files/nandan-nilekani-sabot-lecture-transcript-technology-leapfrog-development.pdf (accessed 9 October 2017).

Nussbaum M and Sen A (1993) *The Quality of Life*. Oxford: Oxford University Press.

Nussbaum MC (2006) *Frontiers of Justice: Disability, Nationality, and Species Membership*. Cambridge, MA: Harvard University Press.

O'Neil C (2016) *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown Publishing Group.

Panikkar R (1982) Is the notion of human rights a western concept? *Diogenes* 30(120): 75–102.

Priya S and Priya A (2016) Even in Delhi, basing PDS on Aadhaar is denying many the right to food. *The Wire*. Available at: https://thewire.in/75359/right-to-food-how-aadhaar-in-pds-is-denying-rights/ (accessed 8 February 2017).

Ram A (2015) Smartphones bring solace and aid to desperate refugees. *Wired*. Available at: https://www.wired.com/2015/12/smartphone-syrian-refugee-crisis/ (accessed 22 June 2017).

Raymond NA (2016) Beyond ''do no harm'' and individual consent: Reckoning with the emerging ethical challenges of civil society's use of data. In: Taylor L, Floridi L and van der Sloot B (eds) *Group Privacy: New Challenges of Data Technologies*. Springer International, pp. 67–82.

Sabhikhi IA (2016) Aadhaar in MGNREGA is likely to be hugely disruptive for workers. *The Wire*. Available at: https://thewire.in/102103/aadhaar-mgnrega-errors-corruption/ (accessed 8 February 2017).

Schwartz PM and Solove DJ (2011) The PII problem: Privacy and a new concept of personally identifiable information. *New York University Law Review* 86: 1814.

Sen A (1999) *Development as Freedom*. New York: Random House.

Sen A (2005) Human rights and capabilities. *Journal of Human Development* 6(2): 151–166.

Shearmur R (2015) Dazzled by data: Big Data, the census and urban geography. *Urban Geography* 36(7): 965–968.

Spivak GC (1988) Can the subaltern speak? In: Nelson C and Grossberg L (eds) *Marxism and the Interpretation of Culture*. University of Illinois Press, pp. 271–313.

Taylor L (2016a) From zero to hero: How zero-rating became a debate about human rights. *IEEE Internet Computing* 20(4): 79–83.

Taylor L (2016b) No place to hide? The ethics and analytics of tracking mobility using mobile phone data. *Environment and Planning D: Society and Space* 34(2): 319–336.

Taylor L (2016c) Safety in numbers? Group privacy and big data analytics in the developing world. *Group Privacy: New Challenges of Data Technologies*. Dordrecht: Springer.

Taylor L (2016d) The ethics of big data as a public good: Which public? Whose good? *Philosophical Transactions of the Royal Society A* 374: 1–13.

Taylor L and Broeders D (2015) In the name of development: Power, profit and the datafication of the global South. *Geoforum* 64(4): 229–237.

Thatcher J (2014) Living on fumes: Digital footprints, data fumes, and the limitations of spatial big data. *International Journal of Communication* 8: 19.

Thikkavarapu PP (2016) The Aadhaar bill is yet another legislation that leaves too much power with the government at the centre. *The Caravan*. Available at: http://www.caravanmagazine.in/vantage/aadhaar-bill-another-legistlation-leaves-power-centre (accessed 22 June 2017).

Tribunales Colegiados del Circuito. *Gaceta del Seminario Judicial de la Federación*, Décima Época, Tomo II, Libro 20, July 2015, p.1719, Tesis 11.1o.29 P (10a), Registro 2009626.

Turow J, Hennessy M and Draper N (2015) The tradeoff fallacy: How marketers are misrepresenting American consumers and opening them up to exploitation. Epub ahead of print 2015. DOI: 10.2139/ssrn.2820060. Available at: http://dx.doi.org/10.2139/ssrn.2820060 (accessed 9 October 2017).

UN General Assembly (1948) Universal declaration of human rights. *United Nations*, 1–6.

United Nations (2014) A world that counts: Mobilising the data revolution for sustainable development. New York. Available at: http://www.undatarevolution.org/wp-content/uploads/2014/12/A-World-That-Counts2.pdf (accessed 9 October 2017).

Unwin PTH (2009) *ICT4D: Information and Communication Technology for Development*. Cambridge: Cambridge University Press.

World Economic Forum (2011) Personal data: The emergence of a new asset class. New York. Available at: http://www3.weforum.org/docs/WEF_ITTC_PersonalDataNewAsset_Report_2011.pdf (accessed 9 October 2017).

Yadav A (2016) In Rajasthan, there is ''unrest at the ration shop'' because of error-ridden Aadhaar. *Scroll.in*. Available at: http://scroll.in/article/805909/in-rajasthan-there-is-unrest-at-the-ration-shop-because-of-error-ridden-aadhaar (accessed 9 October 2017).